

ПІДГОТОВКА ДАНИХ ЯК ОДИН З НАЙВАЖЛИВІШИХ ЕТАПІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ПРОЦЕСІВ

Манєва Р.І., Ісаков О.С., Іващенко О.В.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків*

Process Mining (інтелектуальний аналіз процесів) - ряд методів і підходів, призначених для аналізу та удосконалення процесів в інформаційних системах або бізнес-процесів на підставі вивчення журналів подій. Методологія Process Mining дозволяє отримати знання про структуру і поведінку процесу з журналів подій, що створюються інформаційними системами під час функціонування. Беручи до уваги те, що всі сучасні автоматизовані системи володіють журналами подій, то застосування методів інтелектуального аналізу процесів є потужним інструментом для управління ефективністю на сучасному підприємстві. Виділяють такі етапи інтелектуального аналізу: підготовка даних (формування таблиць даних на основі логів системи); обробка даних (побудова схеми процесу на основі обробки таблиць); аналіз опрацьованих даних (перевірка відповідності процесів та побудова схеми окремих підпроцесів); формування рекомендацій (вдосконалення підпроцесів та формування нових вимог до бізнес процесу).

В рамках етапу підготовки даних необхідно отримати таблицю з даними певного виду, яка буде містити інформацію про процес. Одним із видів таких таблиць є вхідний журнал (лог). Кожен рядок у вхідному журналі відповідає окремій події. У свою чергу, кожна подія несе в собі інформацію про випадок, який створив її, про виконану в її рамках діяльність та час її реєстрації. Подібні журнали подій можна розглядати як сукупність випадків, а окремі випадки – як послідовності подій, що посилаються на них. Виділимо основні атрибути подій в логах: ідентифікатор випадку (caseid); діяльність (activity name); відмітка часу (time stamp); ресурс (resource); інші дані (other data).

Варто зазначити, що на етапі підготовки перед дослідником виникають проблеми повноти даних, отриманих з реальної системи, заповнення пропущених даних та перевірки достовірності даних, або їх статистичної вірогідності, правдивості. Залежно від типу порожнього атрибута можна вибрати управляюче рішення, використовуючи машинне навчання, а саме методи кластеризації, при цьому всі події розбивають на категорії і на підставі отриманих даних, використовуючи метод k-середніх, визначається категорія невідомої події, до якої вона може належати. При низькій вірогідності достовірності даних події або будь-яких інших проблемах, подія видаляється. Проблема неповноти даних вирішується поетапно, при можливості доповнюється, в іншому випадку задача вирішується видаленням всього кортежу події, а самі дані вважаються помилковими. Таким чином, підготовка даних для інтелектуального аналізу процесів може розглядатися як окрема науково-практична задача, вирішення якої забезпечить якість інформації для подальшого прийняття рішень.